**Classification of genes**

We classify genes according to the values of three parameters:

* Replication time cutoff (RTC): A gene is labeled as “early replicated” in a given cell line if its replication time is before RTC. Inversely, it is labeled as “late replicated” in a given cell line if its replication time is after RTC.
* Stringency (PCT): A gene is labeled “expressed when early replicated” if it is expressed in at least PCT percent of the cell lines in which it is labeled as early replicated.
* Switch cutoff (SC): A gene is labeled as “switching” if the difference between its earliest and latest replication times across all cell lines is greater than SC.

Given the values of RTC, PCT and SC, we classify all genes into three classes: E, C and O as follows:

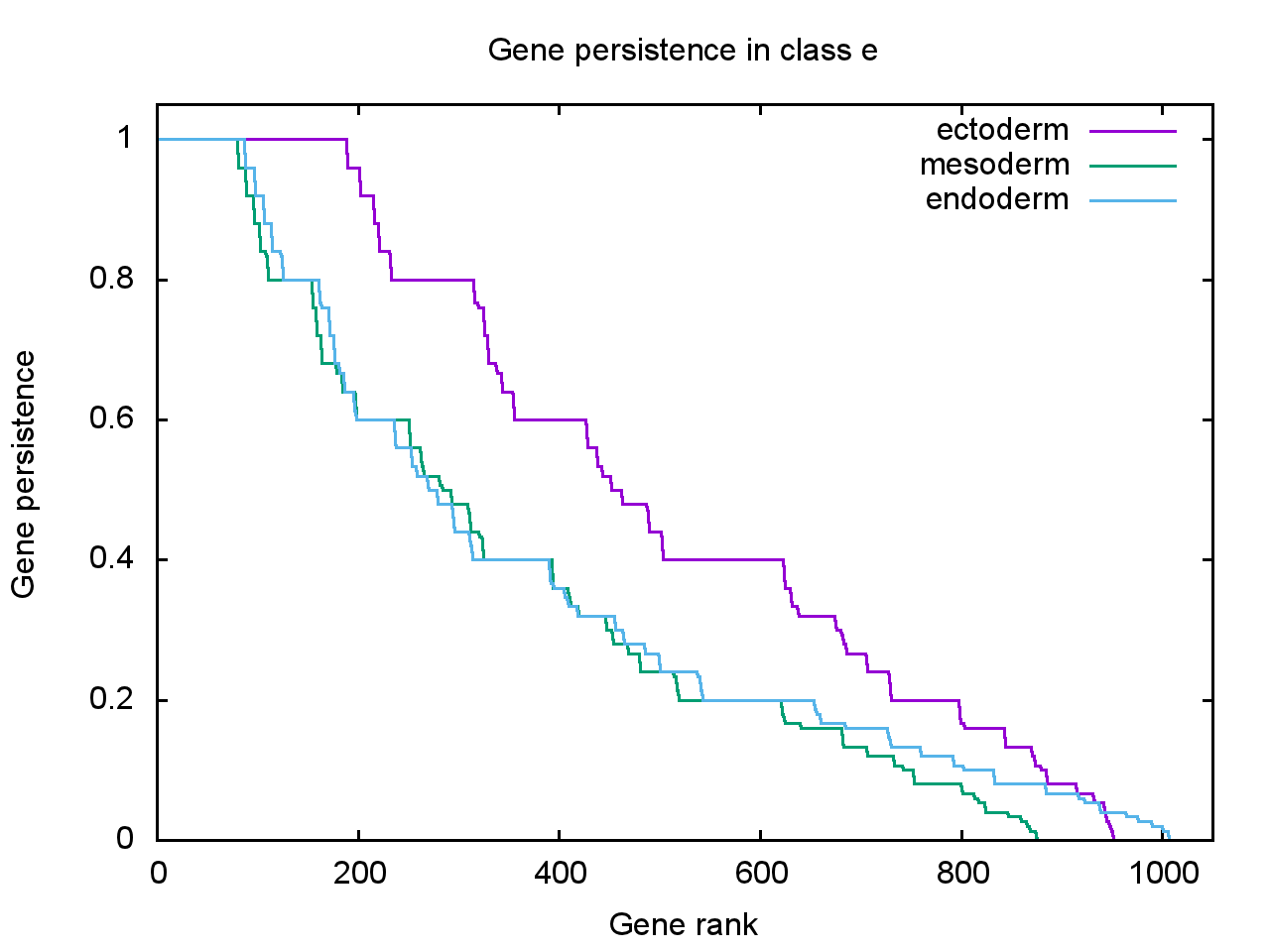
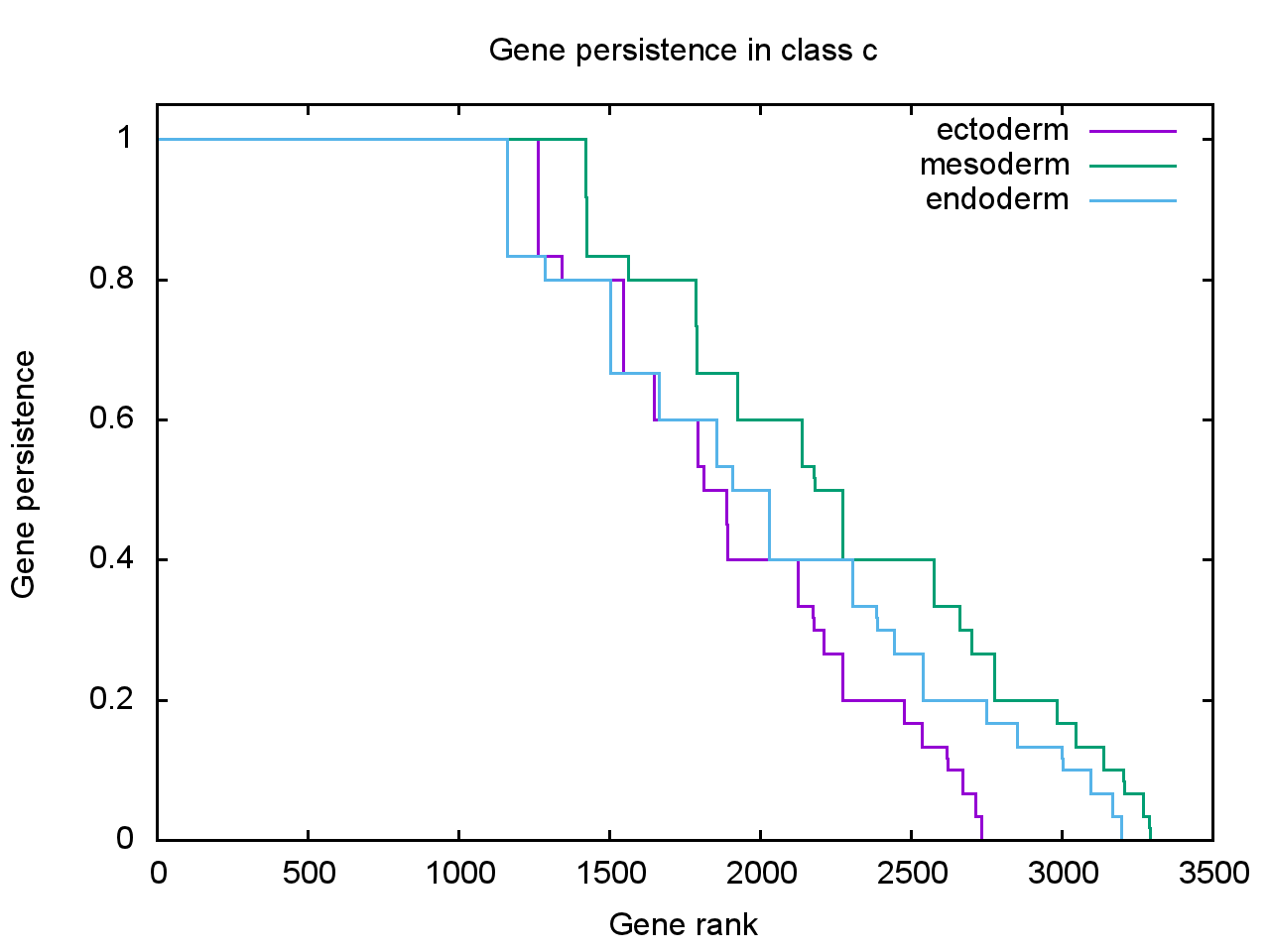
* A gene is assigned to class E if it is 1) switching, 2) early replicated in at least one cell line, 3) late replicated in at least one cell line, 4) expressed when early replicated, and 5) never expressed in any of the cell lines where it is late replicated.
* A gene is assigned to class C if it is 1) switching, 2) early replicated and expressed in at least one cell line, and 3) late replicated and expressed in at least one cell line.
* A gene is assigned to class O if it fits the criteria of neither class E nor class O.

We vary the values of RTC from -0.2 to 0.2, PCT from 10 to 50, and SC from 0.5 to 1, and perform the classification for every combination of the three parameter values. We also perform a germ-layer specific version of this classification by performing it once for every germ layer using the cell lines that belong to this germ layer only. As a result, we have four versions of this classification: Ectoderm, Endoderm, Mesoderm and all combined.

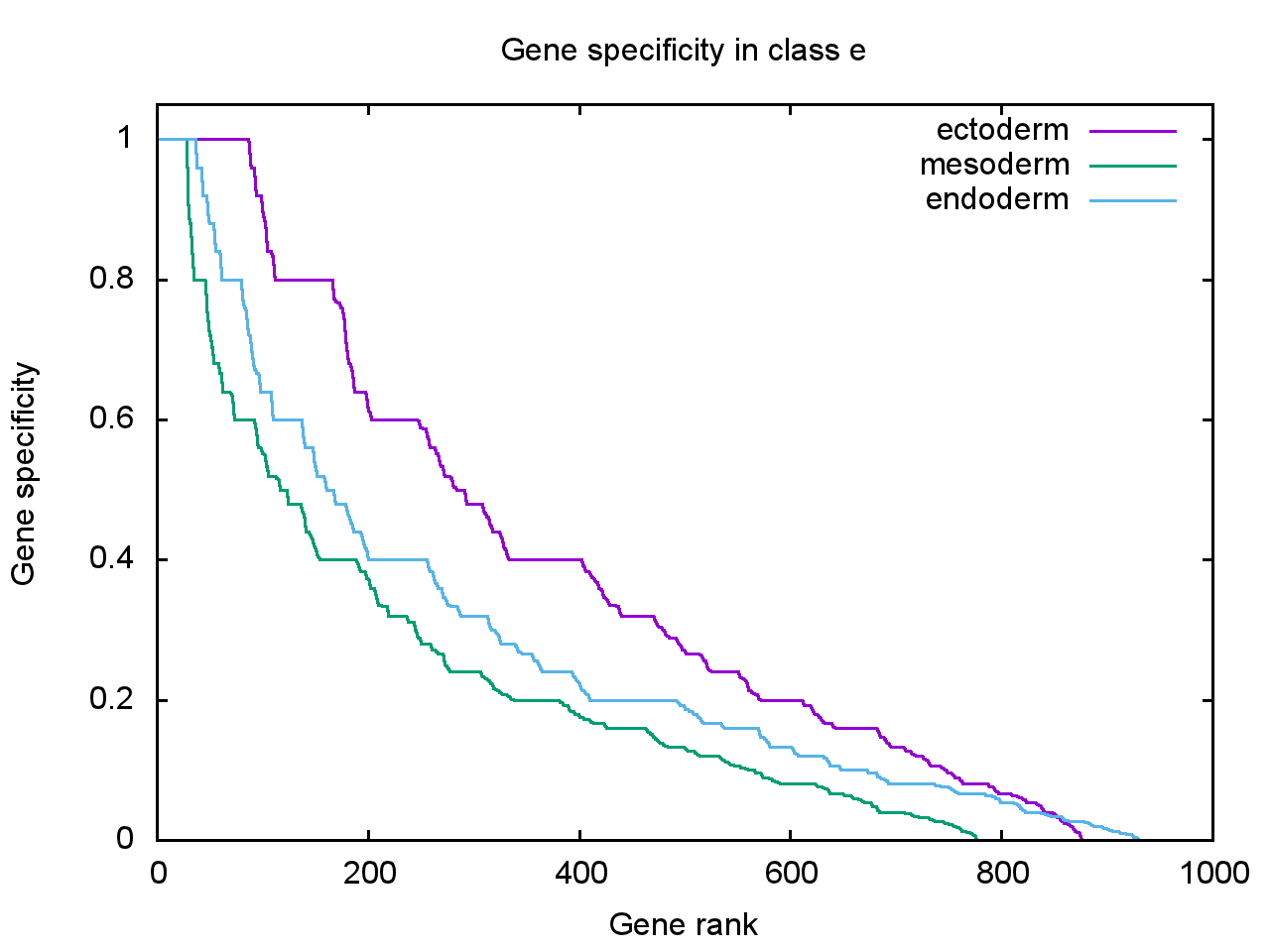
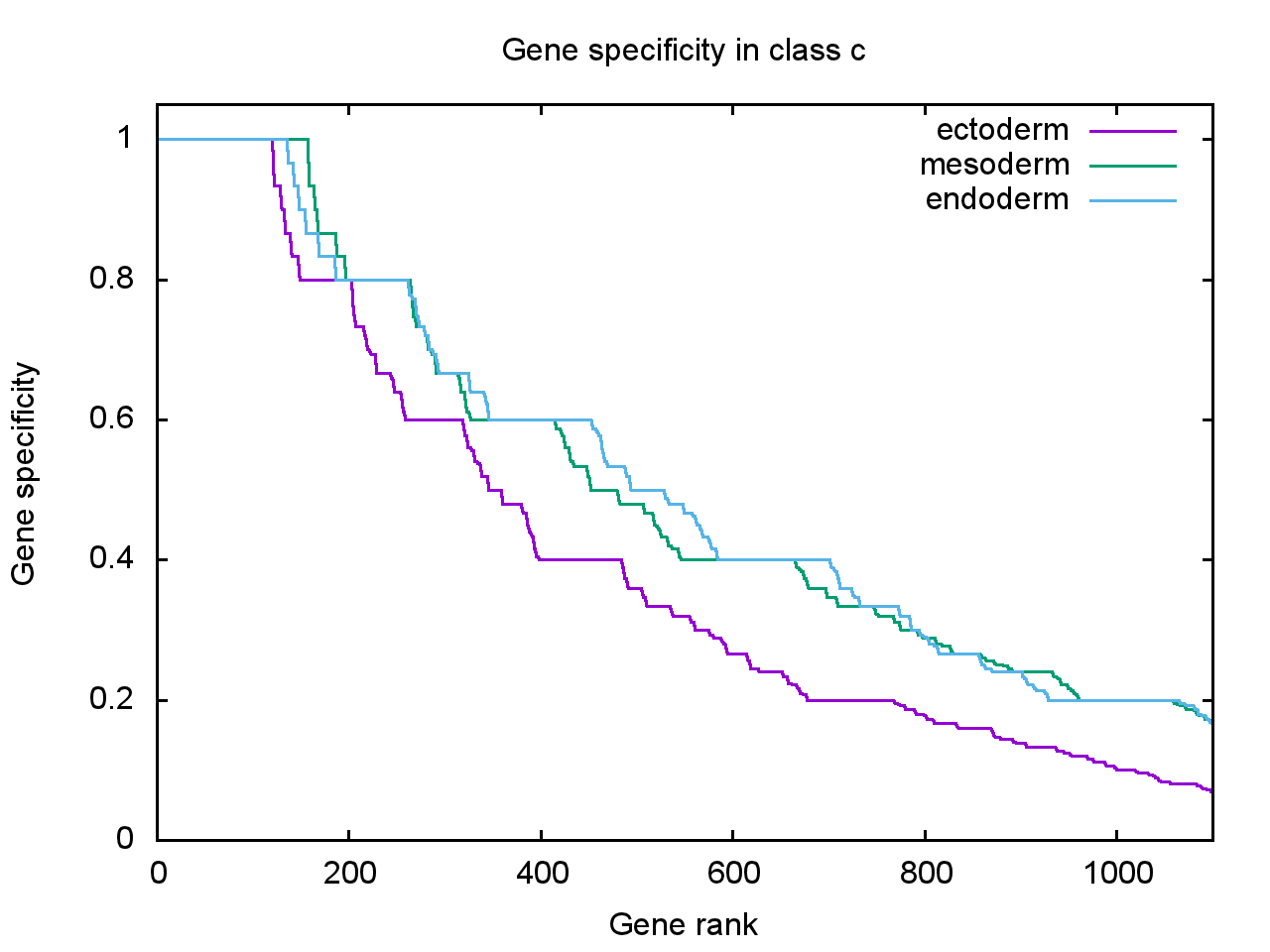
For each of the four versions of the classification, we want to observe the genes that fall in classes E and C most of the time. To achieve this, we compute the “persistence” of a given gene to a given class as the fraction of parameter values for which this gene fell in this class. For germ-layer specific analysis, we also want to observe the genes that fall in class E and C specifically in this germ layer and not in the others. To achieve this, we compute the “specificity” of a given gene to a given class in a given germ layer as the fraction of parameter values for which this gene fell in this class in this germ layer and not in the others.

**Regulation density between classes**

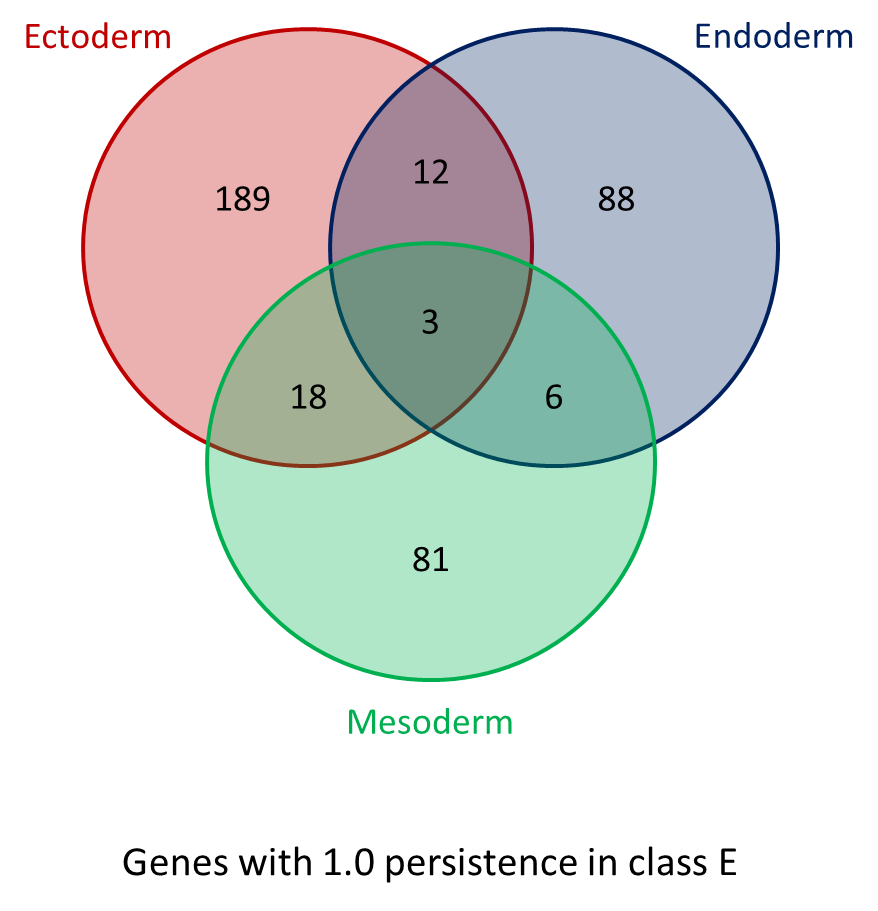
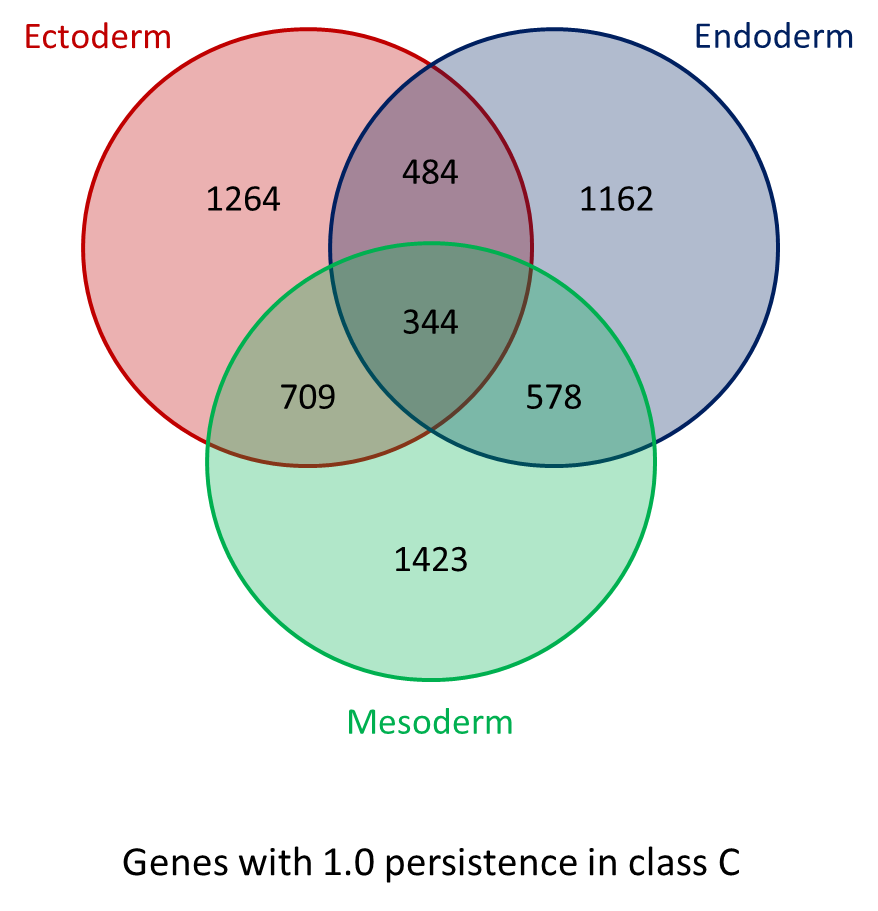
We seek to observe the regulatory relationships among E, C and O classes. We use the transcription factor regulatory networks from Neph et al. [citation] to count the regulatory edges between pairs of classes. For each germ layer, we construct one network by combining all regulatory networks that belong to this germ layer’s cell lines. We also construct a universal network for the combined analysis by combining the three germ-layer specific networks. For each germ layer network as well as the combined network, we compute the “regulation density” from a given class A to another class B as follows. We count the number of edges directed from a gene in class A to a gene class B, and divide it by the total number of possible edges that can be drawn from genes in class A to class B.



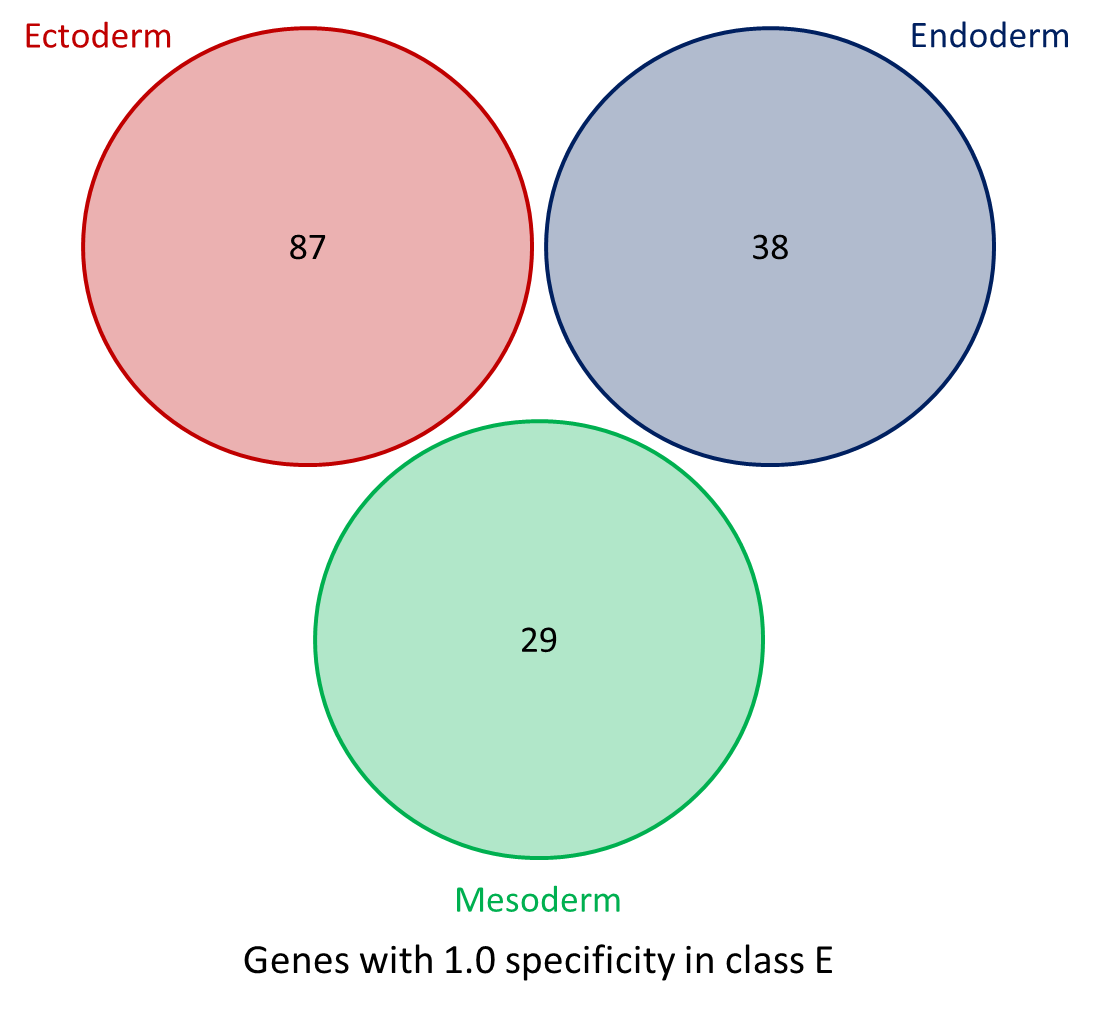
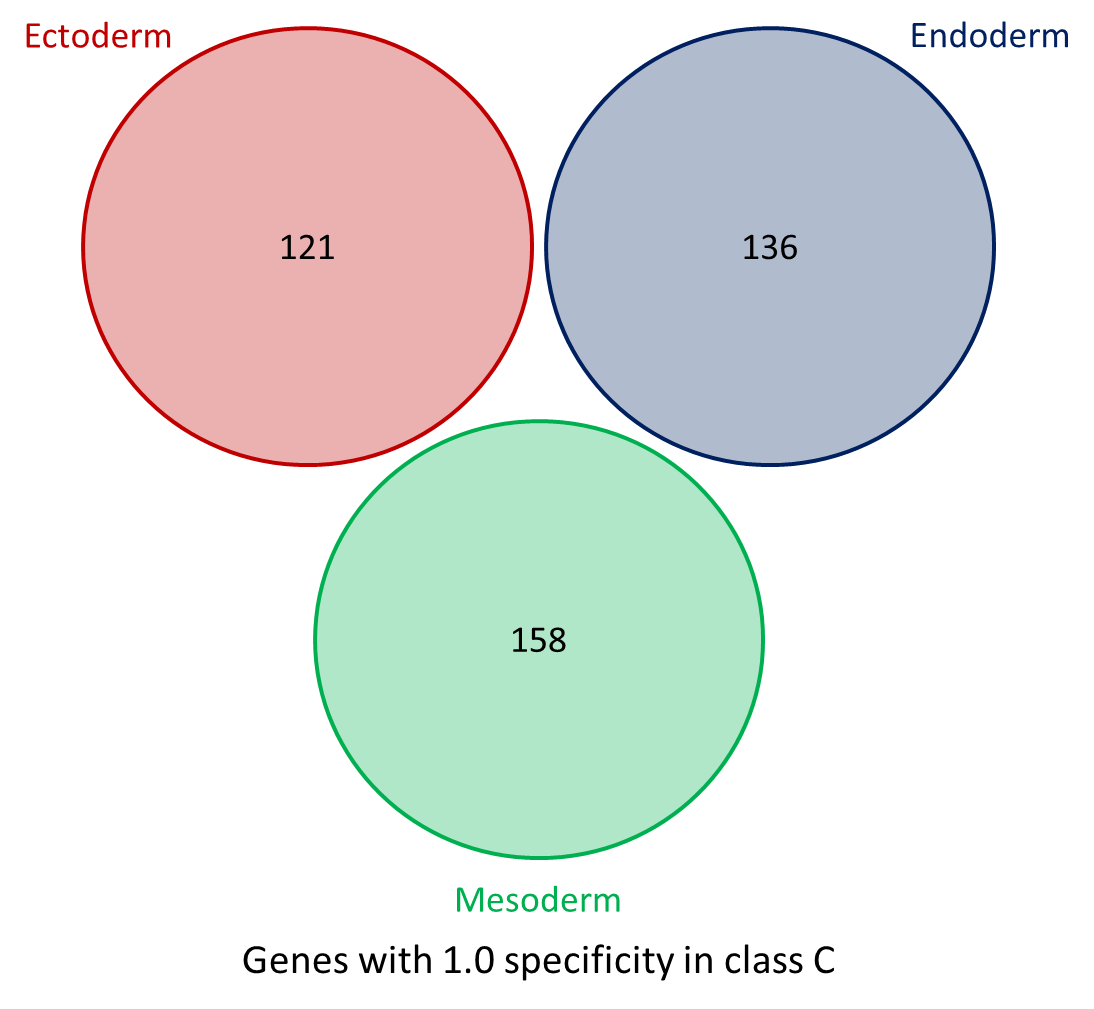
Distribution of gene persistence values in classes C and E across the three germ layers. Gene persistence to a given class is computed as the fraction of parameter values in which this gene falls in this class. X-axis: all genes ranked in descending order of persistence (zero persistence genes are truncated). Y-axis: Gene persistence.



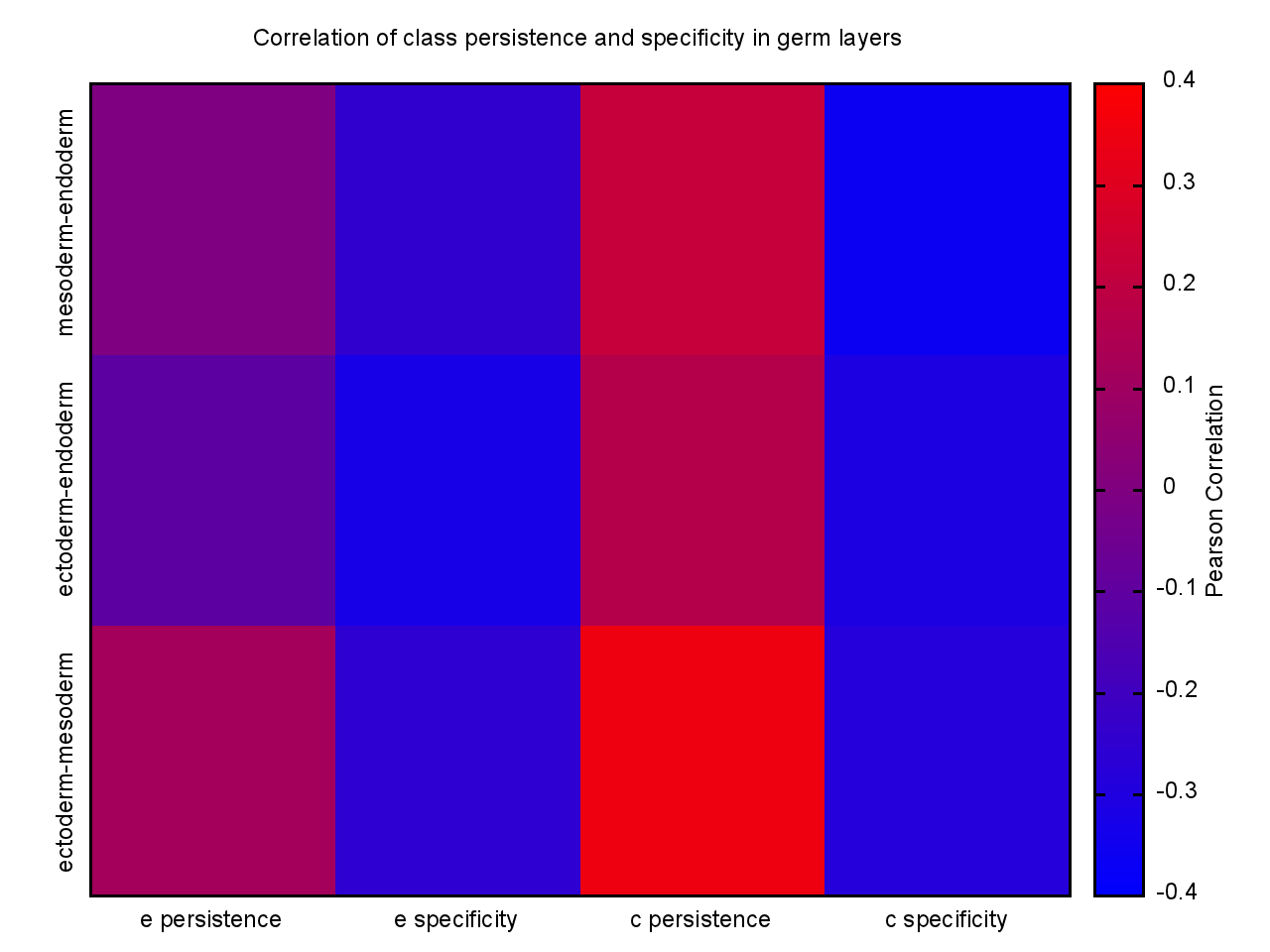
Distribution of gene specificity values in classes C and E across the three germ layers. Gene specificity to a given class in a given germ layer is computed as the fraction of parameter values in which this gene falls in this class in this layer and not in the other layers. X-axis: all genes ranked in descending order of specificity (zero specificity genes are truncated). Y-axis: Gene specificity.



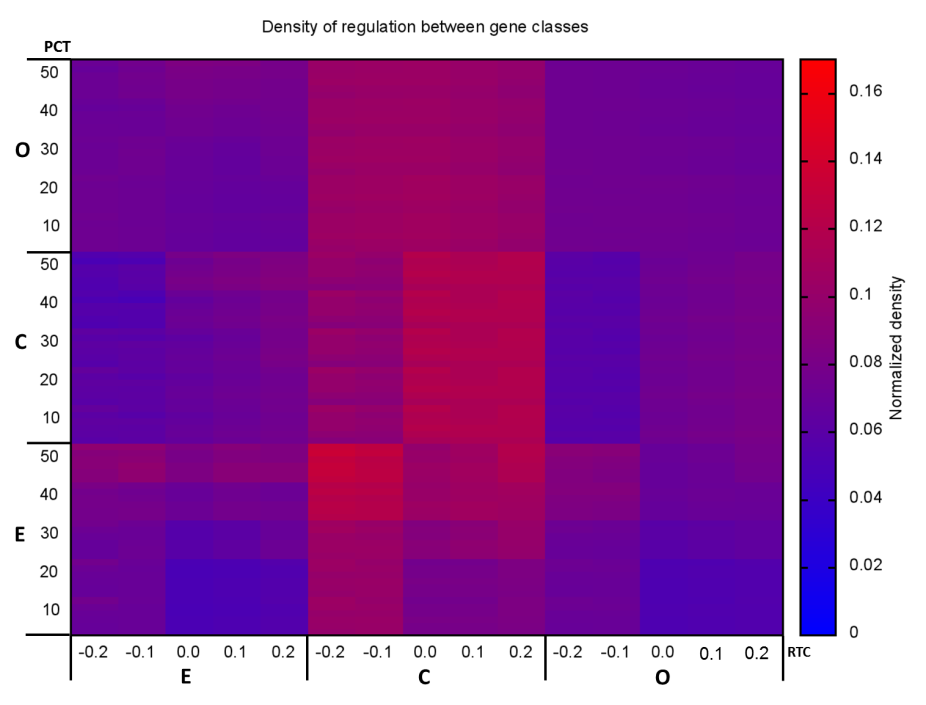
Number of genes with 100% persistence to C and E in each germ layer and their overlap.



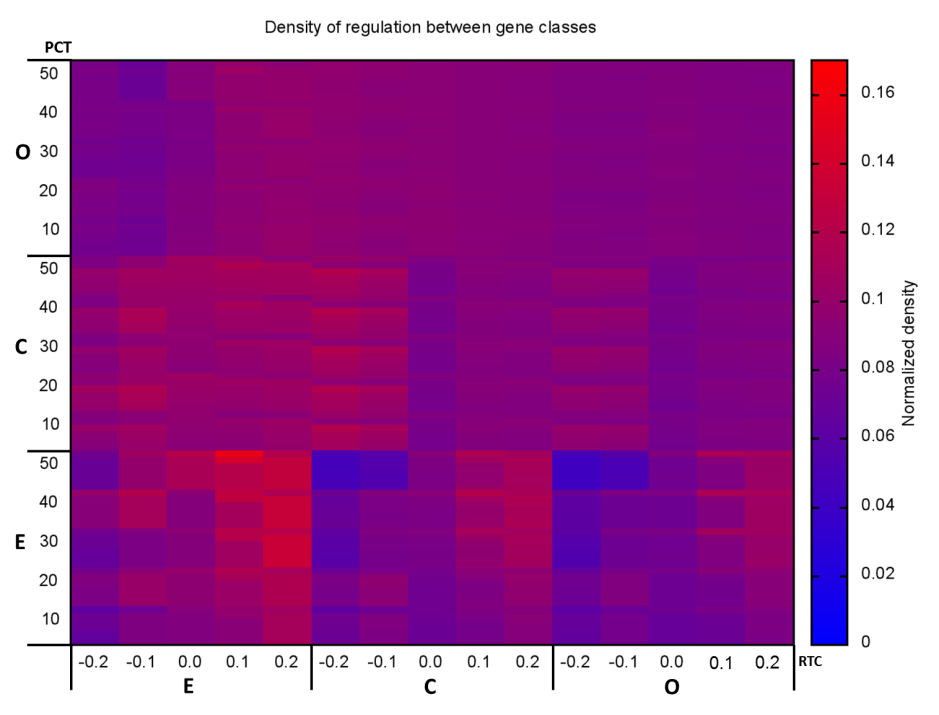
Number of genes with 100% specificity to C and E in each germ layer.



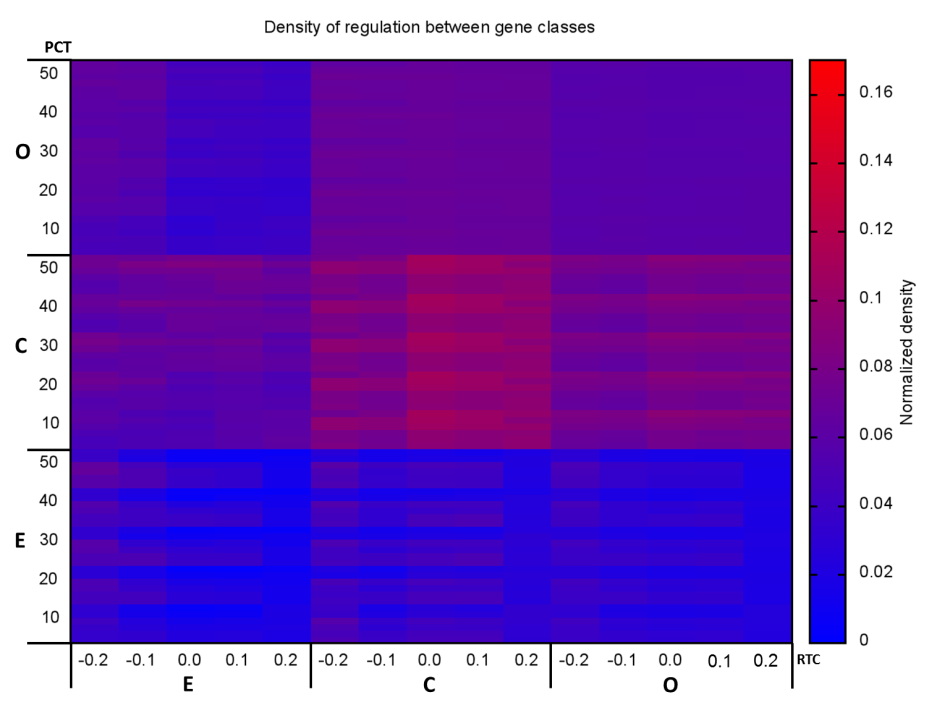
Correlation between germ layers with respect to gene persistence and specificity to classes C and E. Genes that have zero persistence/specificity in all germ layers are ignored. Persistence and specificity values of the remaining genes are sorted according to the alphabetical order of the gene names and stored in a vector. Pearson correlation coefficient is computed between every pair of germ layers with respect to their C and E persistence and specificity vectors.



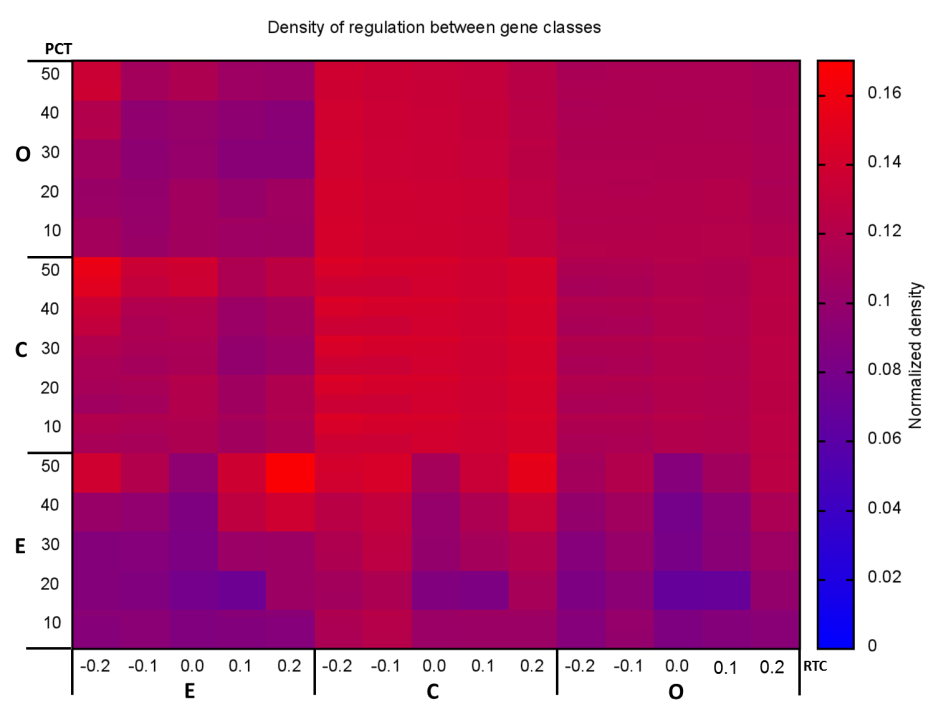
1. Ectoderm



1. Mesoderm



1. Endoderm



1. Combined

Regulation density between pairs of classes using different values of RTC, PCT and SC. Each PCT row is a composite of six rows corresponding to SC values of [0.5 : 1.0]. Regulation density is computed from class x to class y as the number of edges directed from a gene in class x to a gene in class y, divided by the total number of possible edges that can be drawn from a gene in class x to a gene in class y.